Forecasting Exchange Rate Volatility: 
The Superior Performance of Conditional Combinations of Time 
Series and Option Implied Forecasts.*

Guillermo Benavides† Carlos Capistrán‡
Banco de México Banco de México

December 7, 2009

Abstract

This paper provides empirical evidence that combinations of option implied and time se-
ries volatility forecasts that are conditional on current information are statistically superior to
individual models, unconditional combinations, and hybrid forecasts. Superior forecasting per-
formance is achieved by both, taking into account the conditional expected performance of each
model given current information, and combining individual forecasts. The method used in this
paper to produce conditional combinations extends the application of conditional predictive
ability tests to select forecast combinations. The application is for volatility forecasts of the
Mexican Peso–US Dollar exchange rate, where realized volatility calculated using intraday data
is used as a proxy for the (latent) daily volatility.

KEYWORDS: Composite Forecasts, Forecast Evaluation, GARCH, Implied volatility, Mex-
ican Peso - U.S. Dollar Exchange Rate, Regime-Switching.

JEL Classifications: C22, C52, C53, G10.

---

*We thank Alejandro Díaz de León, Antonio E. Noriega, Carla Ysusi, Carlos Muñoz Hink and seminar participants
at the 2008 Latin American Meeting of the Econometric Society at Rio de Janeiro, the XII Meeting of CEMLA’s
Central Bank Researchers’ Network at Banco de España, the 2008 Meeting of the Society of Nonlinear Dynamics
and Econometrics at the Federal Reserve Bank of San Francisco, Banco de México, ITAM, IITESM Campus Cd.
de México, and Universidad del Valle de México for helpful comments. We also thank Antonio Sibaja and Pablo
Bravo for helping us with the exchange rate intraday data. Andrea San Martín, Gabriel López-Moctezuma, Luis
Adrián Muñiz, and Sergio Vargas provided excellent research assistance. The opinions in this paper correspond to
the authors and do not necessarily reflect the point of view of Banco de México.

†Mailing address: Av 5 de Mayo #18, Centro, México, D.F., CP. 06059, México. Tel: (+52)-55-5237-2000 Ext.
3877. Fax: (+52)-55-5237-2559. Email: gbenavid@banxico.org.mx.

‡Mailing address: Av 5 de Mayo #18, Centro, México, D.F., CP. 06059, México. Tel: (+52)-55-5237-2484. Email:
ccapistran@banxico.org.mx.
1 Introduction

Even though several models are used by academics and practitioners to forecast volatility, nowadays there is no consensus about which method is superior in terms of forecasting accuracy (Poon and Granger, 2003; Taylor, 2005; Andersen et al. 2006). The vast majority of models can be classified in two classes: models based on time series, and models based on options (Poon and Granger, 2003). Among the time series models, there are models based on past volatility, such as historical averages of squared price returns, Autoregressive Conditional Heteroskedasticity-type models (ARCH-Type), such as ARCH, GARCH, and EGARCH, and stochastic volatility models. Among the options based volatility models, typically called option implied volatilities (IV), there are the Black-Scholes-type models (Black and Scholes, 1973), the model-free, and those based on hard data on volatility trading. Some authors believe that time series volatility forecasting models are superior because they are specifically designed to capture the persistence of volatility, a salient feature of financial volatility (Canina and Figlewski, 1993; Noh et al. 1994). Others believe that implied volatility is informational superior to forecasts based on time series models because it is the “market’s forecast”, and hence it may be based on a wider information set and also may have a forward looking component (Xu and Taylor, 1995).

In part, the lack of consensus has led researchers to suggest that combining a number of volatility forecasts may be preferable. Patton and Sheppard (2009), and Andersen et al. (2006), among others, highlight the importance of analyzing in more detail composite specifications for volatility forecasting. Becker and Clements (2008) show that combination forecasts of S&P 500 volatility are statistically superior to individual forecasts. Pong et al. (2004) and Benavides (2006) show that combinations of backward-looking and forward-looking forecasts can also be successfully used to forecast exchange rate volatility. This is intuitively appealing given that the volatility obtained from forward-looking methods may have different dynamics than the volatility obtained from backward-looking ones. Thus, combining them could be useful to incorporate features of several forecasting methods in one single forecast, aiming at obtaining a more realistic prediction of the volatility of...
a financial asset. In addition, combining forecasts has had a very respectable record forecasting other economic and financial variables (Timmermann, 2006).

There is some evidence in the combination literature, specifically to forecast macroeconomic variables, that time-varying combination schemes that condition the weights on current and past information may outperform linear combinations (Deutsch et al., 1994; Elliott and Timmermann, 2005; Guidolin and Timmermann, 2009). This time-varying type of forecast combinations may be particularly well suited for forecasting financial volatility due to the observed volatility clustering effect. If the dynamics of high volatility are well captured by a particular method (or set of methods), whereas the dynamics of low volatility are well captured by another method (or set of methods), then a time-varying combination of the forecasts may be the right tool to capitalize on their comparative advantage. This may be true even if one method seems to have an absolute advantage, or an absolute disadvantage.

Two specific time-varying combinations captured our attention, one proposed by Deutsch et al. (1994) with weights that change through time in a discrete manner, and the “Hybrid” forecast of Giacomini and White (2006) (GW), that recursively selects the best forecast. Deutsch et al. (1994) propose the use of switching regressions to estimate the appropriate combination weights, although they do not propose a method to select the variables that determine each regime. GW, in contrast, propose a technique, based on their conditional predictive ability test (CPA), to diagnose if current information can be used to select which forecasting model will be more accurate at a specific future date. GW explore the model-selection implications of adopting a conditional perspective with a simple example of a two-step decision rule that tests for equal performance of the competing forecasts and then –in case of rejection- uses currently available information to select the best forecast for the future date of interest.

In this paper we show that the two-step procedure proposed by GW to select forecasting methods can be extended to select forecast combinations. This results in a conditional combination of the type suggested by Deutsch et al. (1994), but with the advantage that we can test if current information can be used to select the future regime. The extension is simple, as it involves the use of unconditional combinations in the second step.

In order to evaluate our proposed methodology, we first evaluate the forecasting accuracy of some of the most commonly used methods for financial volatility forecasting, as well as combinations
of them, using data on the Mexican peso (MXN)–U.S. dollar (USD) exchange rate. The methods applied in this study are: 1) univariate Generalized Autoregressive Conditional Heteroskedasticity Models (GARCH; Bollerslev, 1986; Taylor, 1986); 2) hard data of implied volatilities from quotes recorded on trading in specific over-the-counter option’s exchange rate deltas; 3) linear combinations of the aforementioned models’ forecasts; and, 4) time-varying combinations, or what we call “conditional combinations”.

Our results indicate that statistically superior out-of-sample accuracy in terms of Mean Squared Forecast Errors (MSFEs) is achieved by conditional combinations of GARCH and (hard-data) IV exchange rate volatility forecasts. These time-varying combinations have weights that vary according to the past level of volatility, based on the fact that when the level of realized volatility is high, IV tends to perform better, whereas when the level of volatility is low, GARCH models tend to be relatively more accurate. Thus, our proposed conditional combinations take into account the comparative advantage of each, backward- and forward-looking volatility forecasting methods.

This study is carried out using MXN–USD intraday exchange rate data to construct a proxy for ex-post volatility, that is then used as a benchmark for forecast evaluation purposes. As shown by Andersen and Bollerslev (1998), the intraday data can be used to form more accurate and meaningful ex-post proxies for the (latent) daily volatility than those calculated using daily data. To the best of our knowledge, high frequency data for this specific exchange rate has not been analyzed anywhere in terms of volatility forecasting.

The layout of this paper is as follows. Section 2 discusses the methodology used to obtain the individual forecasts, to combine them in linear and non-linear ways, and to evaluate forecast performance. The data and our proxies for ex-post realized volatility are presented in Section 3. Section 4 contains the empirical results. Additional exercises are presented in Section 5. Finally, Section 6 concludes.

Evidence about the performance of some of these methods for the case of the volatility of the peso-dollar exchange rate, using daily data, can be found in Benavides (2006).
2 Methodology

2.1 Individual models

All the forecasts produced by the individual models described in this paper are one-step-ahead forecasts (i.e., one-trading-day-ahead).

2.1.1 ARCH-type

The first class of models are the ARCH-type models. It is well documented that ARCH-type models can provide accurate estimates of asset price volatility. This is because these type of models capture the time-varying behavior and clustering commonly observed in the volatility of financial data. The predictive out-of-sample superiority of ARCH-type models has been shown in several studies. For the case of exchange rate volatility, to mention a few, we have the following: Cumby et al. (1993), Guo (1996), and Andersen and Bollerslev (1998), among others. For other type of assets there are mix results. However, most of them show that ARCH-type do better than other time series models.4

In particular, in this paper we apply an univariate GARCH(1,1).5 This parsimonious model was chosen from the ARCH-family given the evidence presented in Hansen and Lunde (2005): In a forecast comparison among 330 ARCH-type models, they find no evidence that a GARCH(1,1) model is outperformed in their analysis of exchange rates.6 Asymmetric volatility models (EGARCH, GJR, QGARCH, among others) were not considered because there is no proven statistical evidence that exchange volatilities have asymmetric volatility (Engle and Ng, 1993).7 Finally, fractionally integrated ARCH models were not analyzed due to an important drawback: in some occasions they produce a time trend in volatility, but time trends are usually not observed in volatility (Granger, 2003).

---


5 ARCH-LM tests following Engle’s (1982) methodology were carried out to corroborate that the series under study has ARCH effects. The results indicate that the series rejected the null in favor of ARCH effects. The procedure was carried out using up to seven lags.

6 A specification search for the lag order was also carried out using information criteria. Akaike Information Criterion (AIC) and Schwartz Bayesian Information Criterion (SIC) were applied. In both cases the selected model was a GARCH(1,1).

7 No asymmetric volatility was found for the data considered for this study. To test for asymmetric volatility two methods were carried out. These were the Engle and Ng (1993) method and an analysis of the cross-correlations of the squared vs. non-squared standardized residuals.
The GARCH(1,1) model is estimated applying the standard procedure, as explained in Bollerslev (1986) and Taylor (1986), but using rolling windows. Two rolling windows of fixed size were used. One contains 756 observations (approximately three years of data), and the other one contains 1526 observations (approximately six years of data). Recursive estimation was not considered because the conditional predictive ability test to be applied later can not be used when the forecasts are obtained using expanding windows (Giacomini and White, 2006). The parameters, including the degrees of freedom of the t-distribution, were estimated using maximum likelihood applying the Marquandt procedure.

2.1.2 Option-implied

It is widely known that implied volatilities from options prices are accurate estimators of the price volatility of their underlying assets (Fleming, 1998; Blair et al. 2001; Taylor, 2005). The forward-looking nature of IV is intuitively appealing and theoretically different from the well-known historical backward-looking conditional volatility ARCH-type models.

Within the academic literature there is evidence that the information content of estimated IV from options could be superior to those estimated with time series approaches (see, among others, Fleming et al. (1995) for futures market indexes; Fleming (1998), Blair et al. (2001) for stocks; Ederington and Guan (2002) for futures options of the S&P 500; and Manfredo et al. (2001) and Benavides (2003) for agricultural commodities). On the other hand, some research papers are skeptical about their out-of-sample forecasting accuracy (see, among others, Day and Lewis, 1992; Canina and Figlewski, 1993; and Lamoureux and Lastrapes, 1993). In terms of forecasting exchange rate volatility, most of the literature has found that IV has both higher accuracy and information content. The evidence is supported by the works of Jorion (1995), Szakmary et al. (2003), and Benavides (2006). Even though there is not a consensus about which method forecasts with higher accuracy, for exchange rate volatility forecasting there is a tendency to observe that IV has better performance compared to its counterparts (Poon and Granger, 2003).

The option IV of an underlying asset is the market’s forecast of the volatility of such asset during

---

8 Nowadays there is no consensus about which method to use in order to find an optimal rolling window size (Pesaran and Timmermann, 2007).

9 Given that these type of models are well documented in the literature, no more details about these are given here. If more details are needed the interested reader can consult Bollerslev et al. (1994). In the present paper these calculations were performed using the UCSD GARCH toolbox for Matlab.
the life of the option. IV is obtained implicitly from the options written on the underlying asset. In the present paper, we use hard-data on IV. Considering that volatility is the only unobservable variable in an option-valuation model, traders in over-the-counter option markets do the trading on volatility-quotes. This latter measurement is hard-data of IVs, which are commonly used by over-the-counter foreign exchange option traders. By a mutual agreement, once the option traders set an IV quote for a transaction it is then substituted into an option-pricing model (usually Garman and Kohlhagen’s 1983 model) in order to determine the option’s price in monetary terms. The price of the option is therefore obtained given the implied volatilities that option traders originally agreed on.

2.2 Linear combinations

Another method used to forecast financial volatility is to combine different forecasting models, resulting in a composite forecast. Each of the models in the composite approach is expected to add significant information to the model as a whole, given that the corresponding forecast errors are not perfectly correlated. It is also well known that the variance of out-of-sample errors can be reduced considerably with composite forecast models (Timmermann, 2006). Composite approaches to forecasting started to be formally considered at least since the late 1960s, with the seminal work of Bates and Granger (1969). Reviews of the now extensive literature on this topic are provided by Clemen (1989) and Timmermann (2006). However, their use for forecasting volatility has been scarce (see Becker and Clements (2008) for a recent application to stock market volatility).

The majority of the research work in the literature about composite models suggests the application of linear combinations. Granger and Ramanathan (1984) consider three regressions:

\( (GR1) \quad \hat{\sigma}^2_{t+1} = \omega_0 + \omega' \hat{\sigma}^2_{t+1|t} + \varepsilon_{t+1} \)
\( (GR2) \quad \hat{\sigma}^2_{t+1} = \omega' \hat{\sigma}^2_{t+1|t} + \varepsilon_{t+1} \) \( (1) \)
\( (GR3) \quad \hat{\sigma}^2_{t+1} = \omega' \hat{\sigma}^2_{t+1|t} + \varepsilon_{t+1}, \text{ s.t. } \omega_i' = 1, \)

where \( \hat{\sigma}^2_{t+1} \) refers to some form of ex-post volatility measure, \( \hat{\sigma}^2_{t+1|t} \) is a \( K \)-vector of ex-ante volatility forecasts, and \( \varepsilon \) is a \( K \times 1 \) vector of ones. The first and second of these regressions can be estimated by standard ordinary least squares (OLS), the only difference being that the second equation
omits an intercept term. The third regression omits an intercept and can be estimated through constrained least squares. The third specification is motivated by an assumption of unbiasedness of the individual forecasts. Imposing that the weights sum to one then guarantees that the combined forecast is also unbiased.\textsuperscript{10} The procedure suggested by Granger and Ramanathan has two desirable characteristics, it yields a combined forecast, which is usually better than either of the individual forecasts, and the method is easy to implement.

In this paper, the three regressions are used. They are estimated recursively (i.e., with an expanding window). In addition, for some variables it has been found that a linear combination using equal weights (i.e., simply averaging the forecasts) yields good results in terms of MSFE (Timmermann, 2006). For this reason, we also consider this linear combination, where the weights, in terms of the above notation, are equal to $\frac{1}{K}$, and $\omega_0 = 0$.

\subsection*{2.3 Conditional combinations}

On some occasions, depending on the particular dynamics of the series to be forecasted, simple combinations as those considered above may not be flexible enough. This is more likely to be true if the performance of each of the forecasts to be combined changes conditional on current and past information.

In this context, Deutsch et al. (1994), among others, pointed out that combinations using time-varying weights may have better performance than simple linear combinations as they may be able to capture some of the non-linearities present in the data. A simple case of time-varying weights is that of regime switching. For two regimes, there can be a set of combination weights for one regime and a different set for the other regime. An important question arises of how to distinguish the regimes, i.e., how to determine when to switch. One approach in the combination literature has been to use observable variables, another has been to use latent variables. The case of observable variables has been studied by Deutsch et al. (1994). The case of latent variables has been studied by Elliott and Timmermann (2005).\textsuperscript{11} In the spirit of Deutsch et al. (1994), in this paper composite forecast models with time-varying weights are estimated using observable variables to determine the regime, hence the present study belongs to the former case.

\begin{footnotesize}
\textsuperscript{10} This specification may not be efficient, however, due to the constraints imposed.
\textsuperscript{11} The latent variable in this context refers to the variable or variables used to determine the regime, and not to the fact that the volatility is itself not observable.
\end{footnotesize}
In particular, Deutsch et al. (1994) propose combining forecasts using changing weights:

\[
\hat{\sigma}_{t+1|t,c}^2 = I(\cdot) \left( \alpha_0 + \alpha \hat{\sigma}_{t+1|t}^2 \right) + (1 - I(\cdot)) \left( \beta_0 + \beta \hat{\sigma}_{t+1|t}^2 \right),
\]

where \( \hat{\sigma}_{t+1|t,c}^2 \) is the combined forecast, and \( I(\cdot) \) is an indicator function. They examine several choices to construct the indicator function, such as past forecast errors or relevant economic variables, in an application to inflation forecasting. Although they succeed in showing that time-varying methods can result in a substantial reduction in MSFE, they do not propose a way to select the variables used to determine the regimes (i.e., to calculate the indicator function), nor do they have a way to test if the time-varying combination has a chance of having a significantly smaller MSFE than linear combinations.

When working with switching regressions, the problem of finding the appropriate variable (or set of variables) to determine a regime is usually encountered. This problem may be even more problematic in a forecasting context, as what is needed is a variable that is able to predict the regime in the future. What we propose in this paper is to use the CPA test of Giacomini and White (2006) to select the appropriate regime. Using this technique, a variable or a set of variables can be tested to see if they can predict which forecast will have a better performance for a particular period in the future. The present paper uses this technique in order to construct an indicator function that can be used to estimate time-varying combinations as in Deutsch et al. (1994).

Giacomini and White (2006) propose a two-step decision rule that uses current information to select the best forecast, between a pair of forecasts, for the future date of interest. The first step performs a CPA test, where the null hypothesis is:

\[
H_0 : E \left[ \left( \hat{\sigma}_{t+1}^2 - \hat{\sigma}_{t+1|t,1}^2 \right)^2 - \left( \hat{\sigma}_{t+1}^2 - \hat{\sigma}_{t+1|t,2}^2 \right)^2 \mid \mathcal{F}_t \right] = 0 \quad t = 1, 2, \ldots,
\]

where \( \hat{\sigma}_{t+1|t,i}^2 \) is the ex-ante volatility forecast produced with model \( i \), \( e_{t+1|t,i} \) is the forecast error from model \( i \), and \( \mathcal{F} \) denotes an information set. From (2), the following orthogonality condition
can be derived:

$$E \left[ h_t \left( e_{t+1|t,1}^2 - e_{t+1|t,2}^2 \right) \right] = 0 \quad \text{for all } \mathcal{F}_t - \text{measurable functions } h_t.$$ 

In this context, $h_t$ is known as the test function. The test can be performed using the (out-of-sample) regression:

$$e_{t+1|t,1}^2 - e_{t+1|t,2}^2 = \delta h_t + \epsilon_t$$

$$H_0 : \delta = 0.$$ 

A rejection is interpreted as implying that $h_t$ contains information to predict which forecast will perform better for the future date of interest. Hence, the first step consists on applying GWs conditional predictive ability test to see if it is possible to find a variable or a set of variables that can predict the future performance of each of the forecasts. The objective is to predict the forecast that will have the smaller loss for the next period.

If the null hypothesis of equal predictive ability is not rejected, then we do not have statistical justification to move on to the second step in the Giacomini and White (2006) procedure. In this case, both forecasting models would have equal predictive accuracy, conditional on the test function used. In particular, if a constant is used as the test function, both forecast have equal unconditional predictive ability and hence, any of them can be used (although a linear combination may work better than each individual forecast).

In case of a rejection, GW’s second step entails the following decision rule: use $\tilde{\sigma}_{t+1,t,2}^2$ if $\hat{d}_{t+1|t} \geq 0$ and use $\tilde{\sigma}_{t+1,t,1}^2$ if $\hat{d}_{t+1|t} < 0$, where $d_{t+1|t} = e_{t+1|t,1}^2 - e_{t+1|t,2}^2$, and $\hat{d}_{t+1|t} = \tilde{\delta} h_t$ is the predicted loss differential. Hence, the second step uses the information from the CPA tests in order to generate an indicator function that can be used to distinguish between two regimes. When the loss differential points out that one forecast will have a better performance in the future, the indicator function selects that forecast. This is a particular type of combination with changing weights where the weights are either zero or one, and there is a variable or set of variables that are used to select the regime.

In this paper, we extend the second step in GW, and propose the following decision rule: use
\[ \alpha_0 + \alpha_1 \hat{\sigma}_{t+1,t,1}^2 + \alpha_2 \hat{\sigma}_{t+1,t,2}^2 \text{ if } \hat{d}_{t+1|t} \geq 0 \text{ and use } \beta_0 + \beta_1 \hat{\sigma}_{t+1,t,1}^2 + \beta_2 \hat{\sigma}_{t+1,t,2}^2 \text{ if } \hat{d}_{t+1|t} < 0. \]

In this context, we would expect \( \hat{\sigma}_{t+1,t,2}^2 \) to receive relatively more weight when \( \hat{d}_{t+1|t} \geq 0 \) and \( \hat{\sigma}_{t+1,t,1}^2 \) to receive relatively more weight when \( \hat{d}_{t+1|t} < 0 \). Hence, we do not go all the way, as GW, to completely switch from one forecast to the other, but maintain the advantages of combinations within regimes.

Once we obtain the indicator function from GW’s first step, we estimate the time-varying combination using OLS on the following regression:

\[
\hat{\sigma}_{t+1}^2 = I(\hat{d}_{t+1|t} \geq 0) \left( \alpha_0 + \alpha_1 \hat{\sigma}_{t+1,t,1}^2 + \alpha_2 \hat{\sigma}_{t+1,t,2}^2 \right) + \left( 1 - I(\hat{d}_{t+1|t} > 0) \right) \left( \beta_0 + \beta_1 \hat{\sigma}_{t+1,t,1}^2 + \beta_2 \hat{\sigma}_{t+1,t,2}^2 \right) + \varepsilon_{t+1}. \tag{3}
\]

### 2.4 Evaluation of forecasting performance

The metric employed in the present paper to test for predictive accuracy is MSFE, calculated for each type of forecast \( i \) as:

\[
MSFE_i = P \sum_{t=1}^{P} \left( \hat{\sigma}_t^2 - \sigma_{t-1,i}^2 \right)^2,
\]

where \( P \) is equal to the number of out-of-sample observations. The choice of loss function considers its simplicity, its generalized use, and that it gives consistent model ranking for commonly used volatility proxies (Patton, 2006).\footnote{Patton (2006) derives necessary and sufficient conditions on the functional form of the loss for the ranking of volatility forecasts to be robust to the presence of noise in the volatility proxy. He shows that MSE loss is robust. Realized volatility is one of the volatility proxies used in Patton (2006) to derive this result.}

The method with the smaller MSFE is considered the most accurate volatility forecasting method. In order to compare the statistical significance of the MSFEs we use Diebold and Mariano’s (1995) test (the Diebold-Mariano-West test, West (1996)).

In order to assess the quality of the ex-ante volatility forecasts, we need an ex-post measure of volatility to use as a benchmark (e.g., to calculate forecast errors and as the dependent variable in combination regressions). A measure typically used in the literature of volatility forecast evaluation is the squared daily return. However, recent research has shown that daily squared returns used as a realized volatility measure are a noisy proxy for that day’s true (latent) volatility (Taylor, 2005; Andersen et al., 2006).

Instead, several papers have emphasized the benefits of using intraday data to calculate a proxy for financial volatility. Realized volatility constructed with intraday data has been proven to be a less noisy proxy for the (latent) daily volatility (Andersen and Bollerslev, 1998).
data is based on the idea that, during a period of time, volatility can be estimated more efficiently as the frequency of returns increases, providing that the intraperiod returns are uncorrelated to each other. However, a trade-off exists, since having the most frequent intraday quotes, say one minute, could increase the bias due to market microstructure effects (e.g., noise from the bid-ask spread, discreteness of the price grid).

The realized variance using intraday data can be calculated as:

\[
RV_t = \frac{1}{\Delta} \sum_{j=1}^{1/\Delta} \left[ p_{t+j\Delta} - p_{t+(j-1)\Delta} \right]^2,
\]

for \( \Delta \) small, positive, and with \( 1/\Delta \gg 1 \). Notice that with \( \Delta = 1 \) the above measure coincides with the squared daily return. Andersen and Bollerslev (1998) suggest that fixing \( \Delta \) at 5 minutes is the frequency that favorably trade-offs the bias and inconsistency induced by microstructure noise and the efficiency achieved by using higher frequencies, and this is probably the most popular choice nowadays.

For the rest of this paper, we take the realized volatility calculated as the realized variance using intraday data at 5 minutes intervals as our ex-post proxy for volatility. Therefore, the evaluation of the forecasting performance of different models is made with respect to this benchmark. However, the exchange rate that we use is not traded as actively as the exchange rates on which the evidence for using 5 minutes intervals is based on. Therefore, we will also report a robustness exercise using as a benchmark the squared daily return.

### 3 Data

#### 3.1 Daily and intradaily returns

The daily data for the spot exchange rate MXN–USD consists of daily spot prices obtained from Banco de México’s web page. These are daily averages of quotes offered by major Mexican banks and other financial intermediaries. The sample period is from January 2nd, 1998 to December 31st, 2007 for a total of 2,499 observations. The intraday data are realizations of the MXN–USD exchange rate with a frequency of 5 minutes (288 observations each day). The transactions were

---

13 Banco de México’s is Mexico’s Central Bank, with web page: [http://www.banxico.org.mx](http://www.banxico.org.mx)
carried out through the Reuters electronic platform. The exact reference is Reuters Matching (RIC: MXN=D2). Within each day, we considered transactions from the 96th observation to the 240th observation. This interval was chosen given that almost 95 percent of the transactions fall within this range. Weekends and holidays were excluded due to significant infrequent trading. The sample size for intraday data is from January 2nd, 2004 to December 31st, 2007. The sample size was chosen considering data availability.\footnote{There were 10 missing trading days in our data base (the first two weeks in May 2004). For those days the daily quotes were used instead (as published by Banco de México). The missing data represent less than 1% of the total number of observations considered for the out-of-sample evaluations.} The observations are taken at each 5-minute interval or the last observation if there is no observation at the exact time interval.

3.2 Realized volatility

The measures of ex-post volatility that we use as the target variable are calculated using equation (4). Our preferred measure uses the intraday data. But we also calculate realized volatility using daily data. Figure 1 shows both estimates for the period from January 2nd, 2004 to December 31st, 2007. As can be seen, the estimate that uses intraday data at five minute intervals captures relatively well the dynamics of the volatility estimated using daily data. However, the former is less noisy. This reflects Andersen and Bollerslev’s (1998) mathematical result that the intraday estimator has lower variance (as $\Delta \to 0$, the variance decreases in the absence of any noise). Detailed advantages of using intraday data to construct realized volatility are also discussed in Andersen et al. (1999).

3.3 Implied volatility

The hard data on IV is for options 1-week and 3-week time to maturity downloaded from UBS—an international financial institution home based in Switzerland—(the tickers are 1WDNMXNUSDImplied for one-week and 3WDNMXNUSDImplied for 3-weeks). As it is known in the options literature a significant number of currency options are traded in terms of annualized volatility. That means that the annualized implied volatility (IV) of the option is quoted (in terms of annualized standard deviations) and transactions are made about them. There are several reasons why we use one and three weeks IV hard data. As it is pointed out in Taylor (2005: page 407) for short horizon forecasts usually nearest to expiry options are the ones that should be used. However,
nearest IV is particularly noisy when days remaining to expiry are few. It is common practice to select one, two or three weeks otherwise the second-nearest is selected. We believe that different maturities in hard data should not be a serious problematic issue given that these are all quotes from at-the-money (ATM) options, which have the same delta. There should be no problems (such as term structure issues) given that we keep consistency having IV with the same delta. In our UBS database there are no IV quotes for options that expire the next-day (to our knowledge there is no historic data available from our database at this very short maturity). Having the same delta these transformations should be close the equivalent (Taylor, 2005).

For our forecasts we calculate the relevant daily volatility from the annualized one. The formula to obtain the scaled implied volatility over $H$ days ($H = 1$ in our case) is given by Taylor (2005, p. 408): $IV = \sqrt{\frac{H}{N}} IV_{\text{annualized}}$. With $N$ the number of days in one year. The usual practice is to consider only trading days given that volatility is significantly lower when markets are closed, therefore, $N = 252$.

4 Empirical results

In the following exercises, the out-of-sample evaluation period is from January 2nd, 2004 to December 31st, 2007 for the individual methods, and from January 2nd, 2006 to December 31st, 2007 for all the methods when the combinations are included among the forecasting methods. In the latter, data for 2004 and 2005 are used to start the estimation of the combination weights.

4.1 Individual models

For the rest of the paper, the abbreviations that we will use are: hard data on implied volatility one-week-to-maturity (iv1week), three-week-to-maturity (iv3week), GARCH rolling using a window of 3 years of daily data (GARCHroll3), and using a window of 6 years (GARCHroll6).

To compare the forecasting performance we use MSFE ratios. We use as the common denominator the MSFE of the GARCH with a rolling window of 6 years given that it is the individual model that performed better in terms of MSFE. The sample used for comparison is January 2, 2004 to December 31, 2007. The MSFE ratios are: GARCHroll3, 1.06; iv1week, 1.33; iv3weeks, 1.19.

Xu and Taylor (1995) obtain similar results from a term structure IV forecast and the nearest or second-nearest IV.
The differences between both GARCHs and the IVs are small, with apparently an advantage for the GARCHs.\footnote{We also evaluated the performance of historical volatility forecasts, using different estimation windows, from 5 days to 252 days, but in all cases they performed poorly.}

Table 1 contains the t-statistics (above the diagonal) and the p-values (below the diagonal) of pair-wise Diebold-Mariano-West tests, where we have used Newey and West’s (1987) heteroskedasticity and autocorrelation consistent covariance matrix estimator. Small p-values can be interpreted as a rejection of the null hypothesis of equal predictive ability. If inference is conducted at the 5% level, it is clear that: (i) both IV hard data show statistically the same performance (p-value is 0.3711); (ii) the GARCH estimated using a rolling window of 6 years is not statistically superior to the one estimated using a rolling window of 3 years (p-value is 0.1164); and (iii) both IV hard data show statistically the same performance than both GARCH models (pairwise), i.e., they have statistically indistinguishable MSFEs, as the DMW predictive ability tests are not able to reject the null of equal predictive accuracy between these forecasts, with the exception of the pair GARCHroll6 and iv1week, for which there is some evidence of superior performance of the ARCH-type forecast.

Unconditionally, there seems to be marginal evidence in favor of the ARCH-type models. More generally, our results confirm that, indeed, it is difficult to choose between forecasts produced using time series methods and those obtained from financial markets (option implieds).

### 4.2 Linear Combinations

The descriptive statistics about the performance of the linear combinations are presented in Table 2. All the combinations include forecasts from the four different methods reported in Table 1. As a reference, the statistics for the GARCH model estimated using 6 years windows are also presented, although in contrast to Table 1, here the analysis is performed using only forecasts for 2006 and 2007, since the data for 2004 and 2005 is used to start the estimation of the combination weights. Looking at the MSFE ratios, all the combinations present MSFE ratios smaller than 1, which indicates that their MSFEs are smaller than that of the best individual model (GARCHroll6) calculated over the same reduced sample, confirming that combining forecasts is a procedure that improves forecasting accuracy. Notice that the combination with equal weights does not produce as good results as the other combinations. The results of the DMW tests applied to the linear
combinations (taken the GARCHroll6 as the benchmark) show that only GR1 and GR2 are able to outperform the best individual model at the 5% significant level.

It is interesting that the best two combinations are GR1, which is the less restrictive of the Granger-Ramanathan regressions, and GR2, the combination that does not include a constant. The latter result is a consequence of: (i) the constant in most of the regressions is close to zero; and (ii) in general, the estimated weights for the GARCH forecasts are usually negative (around -0.5), while the weights estimated for the IVs are usually positive (around 1), which offsets the bias of the individual forecasts. Indeed, the gains from combination arise from a logic similar to the one behind diversification when forming investment portfolios. In this case, because of the negative sign of the weights, the best “portfolio” usually goes short on the GARCHs models. The heterogeneity in the weights also explains the not so good performance of the mean forecasts in this context, and opens the door to the use of time-varying combination schemes that can make better use of it.

4.3 Conditional Combinations

In this section, we first proceed with the conditional evaluation of the individual forecasts (the first step in GW). Then, we use these results to form an indicator function and to estimate the time-varying combinations, following equation (3). Finally, we evaluate the performance of the proposed time-varying composite forecasts.

Table 3 presents the results of the conditional predictive ability test for the individual forecasts, using forecasts from January 2, 2004 to December 30, 2005, since data for 2006 and 2007 are left for out-of-sample evaluation. The Table presents, below the diagonal, the p-values of GW tests using just a constant as an instrument. These tests are similar to the DMW test, in the sense that they can be interpreted as unconditional tests. In this case, at the 1% level there are four null hypotheses that cannot be rejected, those corresponding to the pairs GARCHroll6-GARCHroll3, iv1week-GARCHroll6, iv1week-GARCHroll3, and iv3weeks-GARCHroll3. The interpretation of these results is that, unconditionally, these pairs have MSFEs that are statistically equal. Above the diagonal, the Table presents p-values of CPA tests using as instruments lagged values of the

\footnote{The positive bias in IV forecasts is consistent with the presence of a variance risk premium (e.g., Guo, 1996).}

\footnote{The difference between these results and the DMW results presented in Table 2 are mainly due to the different samples employed.}
realized volatility, as this variable easily suggests itself due to the volatility clustering. In this case, the null of equal conditional predictive accuracy is rejected for only two of these four pairs. For the pair GARCHroll6-iv1week at the 10% level, and for the pair GARCHroll3-iv3weeks at the 1% level. These results indicate that for these pairs the instruments chosen can be used to predict which method will be more accurate the next day.

The important question is how to choose the appropriate rule and the variables to form the indicator function. GW use one lag of the variable to be forecast to predict the loss differential in the future. In accordance to the results presented in Table 3, we use the following regression:

\[ d_{t|t-1} = \gamma_0 + \sum_{i=1}^{k} \gamma_i RV_{t-i} + \sum_{j=1}^{p} \phi_j d_{t-j|t-1-j} + \varepsilon_t, \]  

where \( d_{t|t-1} \) represents the loss differential between each of the two pairs resulting from the application of the CPA tests (e.g., the GARCHroll6 and the iv1week). In contrast to GW, we use lags of the realized volatility as well as lags of the loss differential. This equation is basically a forecasting equation that, using variables known at \( t \), can be used to predict if the loss differential will be positive or negative at \( t + 1 \).

The steps that we followed to produce the conditional combinations are:

1. In the estimation sample, \( t = 1, \ldots, \tau \), we estimate regression (5), selecting the number of lags using AIC.\(^{19}\) The rest of the data are left out for the out-of-sample evaluation.

2. We form a dummy variable as follows:

\[ D_t = \begin{cases} 1 & \text{if } \hat{d}_{t|t-1} \geq 0, \\ 0 & \text{if } \hat{d}_{t|t-1} < 0. \end{cases} \]

3. Next, we use the dummy variable to estimate regression (3) lagged one period:

\[ RV_t = \delta_0 + \delta_1 D_t + \delta_2 \hat{\sigma}_{t|t-1,i}^2 + \delta_3 D_t \hat{\sigma}_{t|t-1,i}^2 + \delta_4 \hat{\sigma}_{t|t-1,j}^2 + \delta_5 D_t \hat{\sigma}_{t|t-1,j}^2 + \varepsilon_t, \]

using data for \( t = 1, \ldots, \tau \), for the pair of forecasts \( i, j \) (e.g., GARCHroll6 and iv1w).

\(^{19}\)The results are robust to the use of SIC instead.
4. To calculate the combined forecast for $\tau + 1$, we need to forecast $\hat{d}_{\tau+1|\tau}$. To do that we use the regression estimated in step 1 above but updating the right-hand variables so that the most recent is at $\tau$.

5. Then, we use $\hat{d}_{\tau+1|\tau}$ to define the value for the dummy $D_{\tau+1}$.

6. Finally, the conditionally combined forecast for $\tau + 1$ is:

$$
\hat{\sigma}^2_{\tau+1|\tau,c}^{(Conditional)} = \hat{\delta}_0 + \hat{\delta}_1 D_{\tau+1} + \hat{\delta}_2 D_{\tau+1|\tau,i} + \hat{\delta}_3 D_{\tau+1|\tau,j} + \hat{\delta}_4 \hat{\sigma}^2_{\tau+1|\tau,i} + \hat{\delta}_5 \hat{\sigma}^2_{\tau+1|\tau,j}.
$$

7. To calculate the forecasts for the next days we repeat steps 1 to 6 but adding one observation at a time, in a recursive manner. For instance, for the next forecast we use the sample $t = 1, ..., \tau + 1$. Always making sure we use observations up to time $t$ to forecast at time $t+1$.

Given the normalization that we are using for the resulting pairs, in which the loss differential is defined as the (squared) error from a GARCH model minus the (squared) error from an IV, when the estimated loss differential is expected to be positive, this reflects periods of high volatility, whereas when the loss differential is expected to be negative, this reflects periods of low volatility. Figure 2 shows that if we separate the returns, and calculate the empirical distribution of the returns when the loss differential is predicted to be positive, and the empirical distribution of the returns when the loss differential is predicted to be negative, positive expected loss differentials are associated with periods of high volatility.

When we use the indicator function to construct hybrid forecasts in the spirit of GW, the rule implies that the GARCHs should be used when this difference is predicted to be negative (i.e., when we predict that the loss function from the GARCH model will be smaller than the loss from the IV), and the IVs hard-data should be used when it is positive. This result is in line with what can be observed in Figure 3. The option implied can follow the realized volatility in periods of high volatility, but it over-predicts during low volatility episodes, in particular with respect to what the GARCH does.

Table 4 presents the MSFE ratios, with respect to the MSFEs of the best individual method, the GARCHroll6, the best linear combination, GR1, the hybrid forecast, and the conditional combination proposed here (all pairwise), calculated over the same sample. Notice that in contrast
to results presented in Table 2, the GR1 combination method reported in Table 4 only combines the corresponding pair of forecasts. The MSFE of the time-varying combinations are smaller than those of any other forecasting method considered in Table 4. In addition, DMW tests confirm that the conditional combinations have MSFEs that are statistically superior to the MSFE of the GARCHroll6, and in one of the two cases is even statistically superior to the MSFE of the best linear combination. The overall finding is that the conditional combinations are superior in terms of MSFE.

One of the reasons why our conditional combination approach may work so well with this data is that the IV forecasts include the volatility risk premium (VRP) (e.g., Bollerslev, et al. 2007). What can be seen in Figure 3 is that the VRP appears to be smaller when the realized volatility is high, and that VRP is a key difference between IV and ARCH-type forecasts. Hence, the conditional combinations give more weight to IV when the VRP is relatively small, and ARCH-type forecast get more weight when the VRP is relatively large.

5 Additional Exercises

5.1 Squared Daily Returns as Proxy for the Daily Volatility

As a robustness check, we repeated the forecasting exercises but using the squared daily returns as our proxy for the (latent) daily volatility. As expected given that this is a noisier measure than the one using intra-day data, the MSFE for all the forecasts and forecast combinations increased. The extra noise certainly makes it more difficult to tell apart the performance of different forecasts. However, in this case we also find that the MSFEs of the conditional combinations are smaller than those of the other methods considered. Table 5 shows the results. One of the pairs used for the combination changed with respect to Table 4. Now the first pair is formed by the GARCHroll3 and the iv1w, instead of the GARCHroll6 and the iv1w. This is a direct result of the CPA tests (not reported) and the fact that now the individual model with the smaller MSFE is the GARCHroll3. Nonetheless, the qualitative results are the same. The MSFE ratios of the conditional combinations are the smaller ratios and the statistical tests of predictive accuracy indicate that they have a very good performance. Furthermore, the fact that the pairs detected by the CPA tests always include a member from each type of forecasts indicates that the conditional combinations are taking into
consideration the comparative advantages of each type.

5.2 AR Models Applied to the Realized Volatility

So far we have limited ourselves to forecasts that can be produced by someone who does not have access to intraday data. Hence, we have only used the realized volatility constructed using intraday data as a proxy for unobservable volatility. However, in recent times, this type of realized volatility has been used to forecast volatility, with encouraging results (Andersen et al. 2006). In order to explore how these models perform with our data and to see if they can also be used with the methods proposed in this paper, we estimated an AR(p) model using the intraday realized volatility. The model is estimated using a rolling window of 370 observations (18 months), with the first sample going from January 1, 2004 to June 30, 2005. The maximum number of lags allowed was 7, and the order was selected in real time using AIC. We label this forecast ARroll.

We compare the forecasts from the ARroll with the GARCH models and with the IVs. The sample used for comparison is July 1, 2005 to December 31, 2007. Using the GARCHroll6 as benchmark and intraday volatility for evaluation, the MSFE ratios are: GARCHroll3, 1.06; iv1week, 1.40; iv3weeks, 1.10; ARroll, 0.91. Clearly, the ARroll is better in terms of MSFE, which corroborates that simple models that take advantage of the intraday data are very good at forecasting volatility (Andersen et al. 2006).

Furthermore, if we apply the GW tests using the sample from January 1, 2005 to December 31, 2006, we find a pair that involves ARroll and that can be used to form conditional combinations. In this case, the pair is formed using the ARroll and the iv1week. For this pair, the p-value using a constant as instrument is 0.2670, hence unconditionally there is no difference in the predictive ability. However, the p-value when one lag of the realized volatility and one lag of the loss differential are used as instruments is 0.0905, indicating that at the 10% level there is evidence of conditional predictability. When we use this pair to form a conditional combination over the sample January 2, 2007 to December 31, 2007, the MSFE ratio between the conditional combination and the ARroll is 0.78, indicating and improvement above 20% with respect to a very good model.

The samples used in the exercise above are arbitrary. However, our point in this sub-section is not to give a full account of the use of intraday data to forecast volatility, but rather to show that our proposed methodology can work well even when this type of forecasts are brought in.
6 Conclusions

The on-going debate regarding which is the most accurate model to forecast volatility of price returns of financial assets has led to a substantial amount of research. Many have compared ARCH-type models against option implied volatilities. Albeit the majority of the literature advocates the use of option implied volatilities as the most accurate alternative to forecast price returns volatilities, there is still no consensus in terms of finding one unique superior model. More recently, following an extensive literature in macroeconomic forecasting, composite forecasts have been shown to be a good way to improve forecasting performance.

In the present paper, the aforementioned volatility forecast models are compared with each other in order to find the most accurate forecasting method for the volatility of the daily spot returns of the Mexican peso – US dollar exchange rate. Intraday exchange rate data with 5 minute time intervals is used to construct our preferred measure of ex-post volatility. The results show that, when comparing single forecasts, there is almost no evidence to reject the null of equal predictive ability between the ARCH-type and (hard data) implied volatilities. Among linear combinations, it is found that, in general, combining results in smaller MSFE with respect to individual forecasts.

However, the best forecast overall is obtained when one ARCH-type forecast and one forecast from the option implieds are combined using time-varying weights. By using past volatility as well as past forecasting performance to predict the regime, a conditional combination that assigns relatively more weight to the option implied during high volatility regimes, and relatively less weight to it during low volatility regimes, has a MSFE that can be 50% lower than the MSFE of the best individual model, and about 25% lower than the MSFE of the best (unconditional) combination, a clear improvement in forecast accuracy.

Future research should look at conditional combinations involving multiple forecasts, and should also explore the conditional combination of forecasts for other variables such as inflation, interest rates, and volatility of other assets.

References


This table reports the results of pair-wise Diebold-Mariano-West tests of equal predictive ability using MSFE as loss function. Above the diagonal we present the Diebold-Mariano statistic, below the diagonal we present the corresponding p-values. In every case, the difference is calculated as the row forecast minus the column forecast.

Abbreviations of the models are as follows: GARCHroll6 and GARCHroll3 denote the GARCH(1,1) model estimated using rolling windows of six and three years respectively and a t-distribution. iv1week and iv3weeks are the hard data option implied volatility with time-to-maturity of one and three weeks, respectively.

<table>
<thead>
<tr>
<th>P-Values</th>
<th>GARCHroll6</th>
<th>GARCHroll3</th>
<th>iv1week</th>
<th>iv3weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCHroll6</td>
<td>–</td>
<td>-1.57</td>
<td>-1.99</td>
<td>-1.84</td>
</tr>
<tr>
<td>GARCHroll3</td>
<td>0.1164</td>
<td>–</td>
<td>-1.61</td>
<td>-1.11</td>
</tr>
<tr>
<td>iv1week</td>
<td>0.0474</td>
<td>0.1079</td>
<td>–</td>
<td>0.89</td>
</tr>
<tr>
<td>iv3weeks</td>
<td>0.0657</td>
<td>0.2679</td>
<td>0.3711</td>
<td>–</td>
</tr>
</tbody>
</table>
Table 2. Evaluation of Forecast Combinations  
(January 2, 2006 - December 31, 2007)  

<table>
<thead>
<tr>
<th>Combination</th>
<th>Mean Error</th>
<th>MSFE Ratio</th>
<th>DMW Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal weights</td>
<td>-4.80E-06</td>
<td>0.86</td>
<td>0.2476</td>
</tr>
<tr>
<td>GR1</td>
<td>-1.94E-06</td>
<td>0.66</td>
<td>0.0271</td>
</tr>
<tr>
<td>GR2</td>
<td>-2.27E-06</td>
<td>0.68</td>
<td>0.0339</td>
</tr>
<tr>
<td>GR3</td>
<td>-4.53E-06</td>
<td>0.80</td>
<td>0.1706</td>
</tr>
<tr>
<td>GARCHroll6</td>
<td>-2.29E-06</td>
<td>1.00</td>
<td>X</td>
</tr>
</tbody>
</table>

This table reports the Mean Error, Mean Square Forecast Error (MSFE) Ratio and p-values of the pair-wise Diebold-Mariano-West tests of the volatility forecasting models for the daily spot returns for the Mexican peso - USD exchange rate. The realized volatility used is the annualized ex-post intraday spot return volatility.

All the combined forecast include: Implied Volatility one week to maturity, Implied Volatility three weeks to maturity, GARCH rolling with three year window and a t-distribution, and GARCH rolling with six year window with a t-distribution.

Abbreviations of the models are as follows: Equal Weights refers to the simple average of the forecasts. GR = Granger-Ramanathan-type model. The first GR combination has a constant and the weights are not restricted, the second lacks a constant, and the third lacks a constant and the weights add to one. GARCHroll6 stands for the GARCH(1,1) model estimated using a rolling window of six years.
<table>
<thead>
<tr>
<th></th>
<th>GARCHroll6</th>
<th>GARCHroll3</th>
<th>iv1week</th>
<th>iv3weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCHroll6</td>
<td>–</td>
<td>0.1380</td>
<td>0.0919</td>
<td>0.0051</td>
</tr>
<tr>
<td>GARCHroll3</td>
<td>0.1915</td>
<td>–</td>
<td>0.2600</td>
<td>0.0075</td>
</tr>
<tr>
<td>iv1week</td>
<td>0.1728</td>
<td>0.5119</td>
<td>–</td>
<td>0.0000</td>
</tr>
<tr>
<td>iv3weeks</td>
<td>0.0013</td>
<td>0.0252</td>
<td>0.0000</td>
<td>–</td>
</tr>
</tbody>
</table>

This table reports p-values of pair-wise Giacomini and White (2006) tests of conditional predictive ability. Below the diagonal the instrument used is just a constant. Above the diagonal the instruments are five lags of the realized volatility. The realized volatility used is the annualized ex-post intraday spot return volatility.

Abbreviations as in Table 1.
<table>
<thead>
<tr>
<th>Forecasts Combination</th>
<th>Mean Error</th>
<th>MSFE Ratio</th>
<th>DMW&lt;sup&gt;a/&lt;/sup&gt;</th>
<th>DMW&lt;sup&gt;b/&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCHroll6 Hybrid</td>
<td>-4.18E-06</td>
<td>1.31</td>
<td>0.1249</td>
<td>0.0502</td>
</tr>
<tr>
<td>GARCHroll6 Conditional</td>
<td>-2.31E-06</td>
<td>0.54</td>
<td>0.0554</td>
<td>0.1796</td>
</tr>
<tr>
<td>iv1w</td>
<td>-2.02E-06</td>
<td>0.69</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>GARCHroll3 Hybrid</td>
<td>-9.64E-07</td>
<td>1.10</td>
<td>0.1009</td>
<td>0.0254</td>
</tr>
<tr>
<td>GARCHroll3 Conditional</td>
<td>-1.63E-06</td>
<td>0.52</td>
<td>0.0408</td>
<td>0.0705</td>
</tr>
<tr>
<td>iv3w</td>
<td>-1.05E-06</td>
<td>0.74</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>GARCHroll6</td>
<td>-2.29E-06</td>
<td>1.00</td>
<td>X</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a/</sup> The benchmark is GARCHroll6
<sup>b/</sup> The benchmark is GR1

This table reports the Mean Error, Mean Square Forecast Error (MSFE) Ratio and p-values of pair-wise Diebold-Mariano-West tests.

Abbreviations of the models are as follows: GARCHroll6, GARCHroll3, iv1week, and iv3week as in Table 1. GR1 as in Table 2.
Table 5. Evaluation of Time-Varying Combinations using Squared Daily Returns as the Proxy for the Daily Volatility
(January 2, 2006 - December 31, 2007)

<table>
<thead>
<tr>
<th>Forecasts</th>
<th>Combination</th>
<th>Mean Error</th>
<th>MSFE Ratio</th>
<th>DMW&lt;sup&gt;a&lt;/sup&gt;</th>
<th>DMW&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCHroll3</td>
<td>Hybrid</td>
<td>-5.11E-06</td>
<td>0.95</td>
<td>0.5562</td>
<td>0.0288</td>
</tr>
<tr>
<td>and</td>
<td>Conditional</td>
<td>-8.61E-07</td>
<td>0.80</td>
<td>0.1014</td>
<td>0.4127</td>
</tr>
<tr>
<td>iv1w</td>
<td>GR1</td>
<td>-8.01E-07</td>
<td>0.85</td>
<td>–</td>
<td>X</td>
</tr>
<tr>
<td>GARCHroll3</td>
<td>Hybrid</td>
<td>-1.42E-07</td>
<td>1.01</td>
<td>0.2134</td>
<td>0.0984</td>
</tr>
<tr>
<td>and</td>
<td>Conditional</td>
<td>9.33E-07</td>
<td>0.83</td>
<td>0.0619</td>
<td>0.1200</td>
</tr>
<tr>
<td>iv3w</td>
<td>GR1</td>
<td>9.96E-07</td>
<td>0.92</td>
<td>–</td>
<td>X</td>
</tr>
<tr>
<td>GARCHroll3</td>
<td></td>
<td>8.72E-08</td>
<td>1.00</td>
<td>X</td>
<td>–</td>
</tr>
</tbody>
</table>

Notes and abbreviations as in Table 4, except that here the benchmark from the individual models is the GARCHroll3.
Figure 1: Exchange rate MXN—USD square daily return and intraday realized volatility.
Figure 2. Indicator function times daily returns January 2, 2006 - December 31, 2007.
Figure 3. Realized volatility and forecasts from GARCH rolling method and hard data option implied 1 week.